Polynomial Certificates for Propositional Classes ⋆

Marta Arias

Center for Computational Learning Systems
Columbia University
New York, NY 10115, USA

Aaron Feigelson

Leydig, Voit & Mayer, Ltd.
Chicago, IL 60601, USA

Roni Khardon *

Department of Computer Science
Tufts University
Medford, MA 02155, USA

Rocco A. Servedio

Department of Computer Science
Columbia University
New York, NY 10027, USA

Abstract

This paper studies the complexity of learning classes of expressions in propositional logic from equivalence queries and membership queries. In particular, we focus on bounding the number of queries that are required to learn the class ignoring computational complexity. This quantity is known to be captured by a combinatorial measure of concept classes known as the certificate complexity. The paper gives new constructions of polynomial size certificates for monotone expressions in conjunctive normal form (CNF), for unate CNF functions where each variable affects the function either positively or negatively but not both ways, and for Horn CNF functions. Lower bounds on certificate size for these classes are derived showing that for some parameter settings the new certificate constructions are optimal. Finally, the paper gives an exponential lower bound on the certificate size for a natural generalization of these classes known as renamable Horn CNF functions, thus implying that the class is not learnable from a polynomial number of queries.

Preprint submitted to Information and Computation *21 February 2008*

1 Introduction

This paper studies the complexity of learning classes of expressions in propositional logic from equivalence queries and membership queries [1]. In this model a learner tries to identify a hidden concept by asking questions. An equivalence query allows the learner to present a hypothesized definition of the concept and ask whether it is correct. If the definition is not correct the learner obtains a counterexample, which can be chosen adversarially. In a membership query the learner presents a potential instance and asks whether it is a member of the concept. The goal of the learner is to identify the concept while using as little resources as possible. Here resources refer both to run time and to the number of queries asked in the process of identifying the concept. Such complexity measures are relevant when we fix a concept class, a set of concepts, from which the hidden concept is chosen. Then a single learner must be able to identify any member of this class in this manner while using bounded resources.

Since its introduction this model has been extensively studied and many classes have been shown to be efficiently learnable. Of particular relevance for the current paper are learning algorithms for monotone expressions in disjunctive normal form (DNF) [2,1], unate DNF expressions [3], and Horn expressions [4,5]. Some results in this model have also been obtained for subclasses of Horn expressions in first order logic but the complexity map there is less clear. Except for a "monotone-like case" [6] the query complexity is either exponential in one of the crucial parameters (e.g. universally quantified variables) [7,8] or the algorithms use additional syntax based oracles [9–11]. It is thus interesting to investigate whether this gap is necessary. Results in [12] show that VC-dimension [13] cannot resolve this question. We therefore need to investigate the certificate complexity [14,15] that more directly captures the query complexity. The current paper takes a first step in this direction by studying the query complexity in the propositional case.

Certificate complexity was introduced by [14,15] (see also [16,17]) who show

⋆ This work has been partly supported by NSF Grant IIS-0099446 (M.A. and R.K.), a Research Semester Fellowship Award from Tufts University (R.K.) and by an NSF Mathematical Sciences Postdoctoral Research Fellowship (R.S.). Work done while A.F. was at the Department of Electrical and Computer Engineering, Northwestern University, while R.S. was at the Division of Engineering and Applied Sciences, Harvard University, and while M.A was at the Department of Computer Science, Tufts University.

* Corresponding author.
Email addresses: marta@cs.columbia.edu (Marta Arias), arf@alumni.northwestern.edu (Aaron Feigelson), roni@cs.tufts.edu (Roni Khardon), rocco@cs.columbia.edu (Rocco A. Servedio).

that a class $\mathcal{C}$ is learnable from a polynomial number of proper equivalence queries (using hypotheses in $\mathcal{C}$) and membership queries if and only if the class $\mathcal{C}$ has polynomial size certificates. This characterization is information theoretic and ignores run time. Certificates have already proved to be a useful tool for studying learnability. For example, conjunctions of unate formulas are learnable with a polynomial number of queries but not learnable in polynomial time unless P=NP [18]. A recent result of [19] shows that DNF expressions require a super-polynomial number of queries even when the hypotheses are larger than the target function by some factor, albeit the factor is small.

This paper establishes lower and upper bounds on certificates for several classes. We give constructions of polynomial certificates for (1) monotone CNF where no variables are negated, (2) unate CNF where by renaming some variables as their negations we get a monotone formula, and (3) Horn CNF where each clause has at most one positive literal. We give certificates in the standard learning model as well as the model of learning from entailment [5] that is studied extensively in Inductive Logic Programming (see e.g. [20]).

The learnability results that follow from these certificate results are weaker than the results in [2,1,3,4] since we obtain query complexity results and the results cited are for time complexity. However, the certificate constructions which we give are different from those implied by these earlier algorithms, so our results may be useful in suggesting new learning algorithms. We also give new lower bounds on certificate size for each of these concept classes. For some parameter settings, our lower bounds imply that our new certificate constructions are exactly optimal.

Finally, we also consider the class of renamable Horn CNF expressions. Note that unate CNF and Horn CNF generalize monotone expressions in two different ways. Renamable Horn expressions combine the two allowing to get a Horn formula after renaming variables. Renamable Horn formulas can be identified in polynomial time and have efficient satisfiability algorithms and are therefore interesting as a knowledge representation [21]. While unate CNF and Horn CNF each have polynomial certificates, we give an exponential lower bound on certificate size for renamable Horn CNF. This proves that the class of renamable Horn CNFs is not learnable in polynomial time from membership and equivalence queries, and answers an open question posed in [22].

We note that recent work of [23] gives strong negative results on learning DNF formulas. More precisely, [23] show that if NP $\neq$ RP then there is no polynomial-time proper PAC learning algorithm for DNF formulas. This result thus provides a *computational* lower bound for proper learning of DNF in the standard PAC setting of learning from random examples only. In contrast, the certificate constructions that we consider have implications for the *information-theoretic* complexity of proper learning algorithms in the frame-

work of exact learning from membership and equivalence queries. Characterizing the query complexity of learning DNF in this model remains an important open question.

2 Preliminaries

We consider families of expressions built from $n \geq 1$ propositional variables. We assume some fixed ordering so that an element of $\{0,1\}^n$ specifies an *assignment* of a truth value to these variables. The *weight* of an assignment is the number of bits that are non-zero.

A *literal* is a variable or its negation. A *term* is a conjunction of literals. A *DNF* expression is a disjunction of terms. A *clause* is a disjunction of literals. A *CNF* expression is a conjunction of clauses. The *DNF size* of a boolean function $f : \{0,1\}^n \to \{0,1\}$, denoted $|f|_{DNF}$, is the minimum number of terms in a DNF representation of f. The *CNF size* of f, $|f|_{CNF}$, is defined analogously. In general, let $\mathcal{R}$ be a representation class for boolean formulas. Then $|f|_{\mathcal{R}}$ is the $\mathcal{R}$-size of a minimal representation for f in $\mathcal{R}$. If $f \notin \mathcal{R}$, we assign $|f|_{\mathcal{R}} = \infty$.

Next, we present some classes of boolean formulas and their properties. In what follows we use the notation $f(x) = 1$ or $x \models f$ interchangeably, where f is a boolean function and x is an assignment. Both stand for classical formula satisfiability. Additionaly, when $x \models f$ we say that x is positive for f and when $x \not\models f$ we say that x is negative for f.

A term t is a *minterm* for a boolean function f if $t \models f$ but $t' \not\models f$ for every other term $t' \subset t$. A DNF representation $t_1 \vee \ldots \vee t_k$ of a boolean function f is *non redundant* if each t_i is a minterm of f and if removing any t_i changes the function. That is, there is an assignment x such that $x \models t_i$ but $x \not\models \vee_{j \neq i} t_j$. Analogously, a CNF representation $c_1 \wedge \ldots \wedge c_k$ is not redundant if each clause is minimal and for all c_i we have $\wedge_{j \neq i} c_j \not\models c_i$.

A *monotone CNF (DNF)* expression is a CNF (DNF) with no negated variables. Semantically, a function is monotone iff:

$$\forall x, y \in \{0,1\}^n : \text{ if } x \leq y \text{ then } f(x) \leq f(y), \tag{1}$$

where $\leq$ between assignments denotes the standard bit-wise comparison relation.

An *anti-monotone CNF (DNF)* expression is a CNF (DNF) where all variables

appear negated. Semantically, a function is anti-monotone iff:

$$\forall x, y \in \{0,1\}^n : \text{ if } x \leq y \text{ then } f(x) \geq f(y). \tag{2}$$

Let $a, x, y \in \{0,1\}^n$ be three assignments. The inequality between assignments $x \leq_a y$ is defined as $x \oplus a \leq y \oplus a$, where $\oplus$ is the bit-wise exclusive OR. Intuitively if $a[i]$, the i'th bit of a, is 0 then we get the normal order on this bit. But if $a[i] = 1$ we use $1 < 0$ for the corresponding variable. We denote $x <_a y$ iff $x \leq_a y$ but $y \not\leq_a x$.

A boolean function f (of arity n) is *unate* iff there exists some assignment a (called an *orientation* for f) such that

$$\forall x, y \in \{0,1\}^n : \text{ if } x \leq_a y \text{ then } f(x) \leq f(y). \tag{3}$$

Equivalently, a variable cannot appear both negated and unnegated in any non-redundant CNF or DNF representation of f. Each variable is either monotone or anti-monotone. It is well known that a unate DNF expression has a unique minimal representation given by the disjunction of its minterms, and similarly the minimal CNF representation is unique.

A *Horn clause* is a clause in which there is at most one positive literal, and a Horn expression is a conjunction of Horn classes. A Horn clause $(\overline{x}_{i_1} \vee \cdots \vee \overline{x}_{i_k} \vee x_{i_{k+1}})$ is easily seen to be equivalent to the implication $x_{i_1} \cdots x_{i_k} \rightarrow x_{i_{k+1}}$; we refer to $x_{i_1} \cdots x_{i_k}$ as the *antecedent* and to $x_{i_{k+1}}$ as the *consequent* of such a clause. Notice that an anti-monotone CNF expression can be seen as a Horn CNF whose clauses have empty consequents. For example, the anti-monotone CNF $(\bar{a} \vee \bar{b}) \wedge (\bar{b} \vee \bar{c})$ is equivalent to the Horn CNF $(ab \rightarrow \texttt{false}) \wedge (bc \rightarrow \texttt{false})$.

Let $x, y \in \{0,1\}^n$ be two assignments. Their intersection $x \cap y$ is the assignment that sets to 1 only those variables that are 1 in both x and y. It is well known that a function is Horn iff

$$\forall x, y \in \{0,1\}^n : \text{ if } x \models f \text{ and } y \models f, \text{ then } x \cap y \models f \tag{4}$$

The original characterization is due to McKinsey [24], although it was stated in a different context and in more general terms. It was further explored by Horn [25]. A proof adapted to our setting can be found e.g. in [26].

Let $a, x, y \in \{0,1\}^n$ be three assignments. Let $a[i]$ be the i-th bit of assignment

a. The unate intersection $x \cap_a y$ is defined as:

$$(x \cap_a y)[i] = \begin{cases} x[i] \wedge y[i] \text{ if } a[i] = 0 \\ x[i] \vee y[i] \text{ otherwise} \end{cases}$$

It is easy to see that this definition is equivalent to $(x \cap_a y)[i] = ((x[i] \oplus a[i]) \cap (y[i] \oplus a[i])) \oplus a[i]$ and that $(x \cap_a y) \leq_a x$ and $(x \cap_a y) \leq_a y$ so that $\leq_a$ and $\cap_a$ behave like their normal counterparts.

We say that a boolean function f (of arity n) is *renamable Horn* if there exists some assignment c such that f_c is Horn, where $f_c(x) = f(x \oplus c)$ for all $x \in \{0,1\}^n$. In other words, the function obtained by taking the complement of variables set to 1 in c is Horn. We call such an assignment c an orientation for f. Equivalently, a function is renamable Horn iff there exists an assignment c such that

$$\forall x, y \in \{0,1\}^n : \text{ if } x \models f \text{ and } y \models f, \text{ then } x \cap_c y \models f. \tag{5}$$

The renamable Horn size of a renamable Horn function f, that is $|f|_{Ren-Horn}$, is the CNF Horn size of $f_c(x)$.

Let $\mathcal{B}$ be any of the classes of propositional expressions defined above; $\mathcal{B}_m$ denotes the subclass of $\mathcal{B}$ whose concepts have size at most m.

The following simple lemma is useful in our constructions:

Lemma 1 *Let c be any clause, t any term, and x, y, a any assignments.*
(1) If $x \not\models c$ and $y \not\models c$ then $x \cap_a y \not\models c$ for any a.
(2) If $x \models t$ and $y \models t$ then $x \cap_a y \models t$ for any a.

Proof: For (1) note that if x and y falsify the clause c then all the variables in c have to share the same value in both x and y. Therefore their "intersection" (w.r.t. any orientation a) does not change the value of these variables in the resulting $x \cap_a y$ implying that that $x \cap_a y$ falsifies c. The same argument can be used to establish (2). □

2.1 Learning with Queries and Certificates

We briefly review the model of exact learning with equivalence queries and membership queries [1]. Before the learning process starts, a concept $c \in \mathcal{B}$ is fixed. We refer to this concept as the *target* concept. The learning algorithm has access to an equivalence oracle and a membership oracle that provide

information about the target concept. In an equivalence query, the learner presents a hypothesis and the oracle answers `Yes` if it is a representation of the target concept. Otherwise, it answers `No` and provides a counterexample, that is, an example $x \in \{0,1\}^n$ where the target and hypothesis disagree. In general the representation of hypotheses is not restricted. However, for *proper learnability* we require that the hypothesis in an equivalence query is represented by a formula in $\mathcal{B}$. In a membership query, the learner presents an example and the oracle answers `Yes` or `No` depending on whether the example presented is a member of the target concept. For any target expression in the concept class the learning algorithm is required to identify the target expression and get a `Yes` answer to an equivalence query.

When concept classes are parametrized by size the notion of proper learnability can be slightly refined. In particular we allow the learning algorithm to learn concepts in $\mathcal{B}_m$ using hypotheses in $\mathcal{B}_{p(m,n)}$ for some polynomial $p()$.

Definition 2 *The* query complexity *of a concept class $\mathcal{B}$, with hypothesis expansion $p(n,m)$, denoted $QC(\mathcal{B},n,m,p(n,m))$, is the minimum number of queries required by any algorithm that learns $\mathcal{B}$ with equivalence queries and membership queries, where the hypotheses are restricted to be in $\mathcal{B}_{p(m,n)}$.*

If $p(n,m)$ is a polynomial and QC is polynomial in n,m then we say that $\mathcal{B}$ is properly learnable *with a polynomial number of queries. If $p(n,m)=m$ and QC is polynomial in n,m then we say that $\mathcal{B}$ is* strongly properly learnable *with a polynomial number of queries.*

Certificates are similarly defined relative to expansions in representation of concepts. Informally, a certificate gives a proof that a function f whose $\mathcal{B}$-size is more than $p(n,m)$ is not in $\mathcal{B}_m$. More formally,

Definition 3 *Let $\mathcal{R}$ be a class of representations defining a boolean concept class $\mathcal{B}$. The class $\mathcal{R}$ has* certificates *of size $q(n,m)$ for representation expansion $p(n,m)$ if for every $n,m>0$ and for every boolean function $f \subseteq \{0,1\}^n$ s.t. $|f|_{\mathcal{R}} > p(m,n)$, there is a set $Q \subseteq \{0,1\}^n$ satisfying the following: (1) $|Q| \leq q(m,n)$ and (2) for every $g \in \mathcal{B}_m$ there is some $x \in Q$ s.t. $g(x) \neq f(x)$. In other words, (2) states that no function in $\mathcal{B}_m$ is consistent with f over Q.*

If $p(\cdot,\cdot)$ and $q(\cdot,\cdot)$ are polynomials then we say that $\mathcal{B}$ has polynomial size certificates. The certificate size of $\mathcal{B}$ for representation expansion $p(n,m)$, denoted $CS(\mathcal{B},n,m,p(n,m))$ is the smallest function $q(n,m)$ which satisfies the above.

We can now state the relation between query complexity and certificates:

Theorem 4 [15,14,16]

$CS(\mathcal{B}, n, m, p(n,m)) \leq QC(\mathcal{B}, n, m, p(n,m)) \leq CS(\mathcal{B}, n, m, p(n,m)) \log(|\mathcal{B}_m|)$

3 Certificates for monotone and unate CNFs

In this section we give constructions of certificates for monotone and unate classes. We present the basic result for the class of anti-monotone CNF so as to make the relation to the certificate for Horn expressions as clear as possible.

Theorem 5 *The class of anti-monotone CNF has polynomial size certificates with* $p(m,n) = m$ *and* $q(m,n) = \min\{(m+1)n, \binom{m+1}{2} + m + 1\}$.

Proof: Fix $m, n > 0$. Fix any $f \subseteq \{0,1\}^n$ s.t. $|f|_{anti-monCNF} > p(m,n) = m$. We proceed by cases.

Case 1. f is not anti-monotone. In this case, there must exist two assignments $x, y \in \{0,1\}^n$ s.t. $x < y$ but $f(x) < f(y)$ (otherwise f would be anti-monotone). Let $Q = \{x, y\}$. Notice that by definition no anti-monotone CNF can be consistent with Q. Moreover, $|Q| = 2 \leq q(m,n)$.

Case 2. f is anti-monotone. Let $c_1 \wedge c_2 \wedge .. \wedge c_m \wedge .. \wedge c_k$ be a minimal representation for f. Notice that $k \geq m+1$ since $|f|_{anti\text{-}monCNF} > p(m,n) = m$.

We give two different constructions for certificates in this case that achieve the two parts in the bound. Define assignment $x^{[c_i]}$ as the assignment that sets to 1 exactly those variables that appear in c_i's antecedent. For example, if $n = 5$ and $c_i = v_3 v_5 \rightarrow \texttt{false}$ then $x^{[c_i]} = 00101$.

Remark 6 Notice that every $x^{[c_i]}$ falsifies c_i (antecedent is satisfied but consequent is `false`) but satisfies every other clause in f. If this were not so, then we would have that some other clause c_j in f is falsified by $x^{[c_i]}$, that is, the antecedent of c_j is true and therefore all variables in c_j appear in c_i as well (i.e. $c_j \subseteq c_i$). This is a contradiction since c_i would be redundant and we are looking at a minimal representation of f.

Let 0_i be the assignment with 0 in position i and 1 elsewhere. For the first construction let $Q_1 = Q_1^+ \cup Q^-$, where

$$Q^- = \left\{ x^{[c_i]} \,\middle|\, 1 \leq i \leq m+1 \right\} \text{ and}$$

$$Q_1^+ = \left\{ x^{[c_i]} \cap 0_j \,\middle|\, 1 \leq i \leq m+1, 1 \leq j \leq n, \text{ and } x^{[c_i]}[j] = 1 \right\}.$$

Notice that $|Q_1| \leq (m+1)n$. By the remark assignments in Q^- falsify f, and since these are maximally negative assignments it is also clear that assignments

in Q_1^+ satisfy f. Any anti-monotone CNF g with at most m clauses will cover two examples $x^{[c_i]}, x^{[c_j]}$ in Q^- with the same clause. Therefore one of the assignments directly below $x^{[c_i]}$ which is in Q_1^+ is also falsified by this clause. So g is not consistent with f over Q_1.

For the second construction let $Q_2 = Q_2^+ \cup Q^-$, where Q^- is defined as above, and

$$Q_2^+ = \left\{ x^{[c_i]} \cap x^{[c_j]} \;\middle|\; 1 \leq i < j \leq m+1 \right\}.$$

Notice that $|Q_2| \leq \binom{m+1}{2} + m + 1$. The assignments in Q_2^+ are positive for f. To see this, suppose some $x^{[c_i]} \cap x^{[c_j]} \in Q_2^+$ falsifies f. Then there is some clause c in f that is falsified by $x^{[c_i]} \cap x^{[c_j]} \in Q_2^+$. That is, all variables in c are set to 1 by $x^{[c_i]} \cap x^{[c_j]} \in Q_2^+$. Therefore, all variables in c are set to 1 by $x^{[c_i]}$ and $x^{[c_j]}$ and they falsify the same clause which is a contradiction by the remark above.

It is left to show that no anti-monotone CNF g s.t. $|g|_{anti\text{-}monCNF} \leq m$ is consistent with f over Q_2. Fix any $g = c'_1 \wedge .. \wedge c'_l$ with $l \leq m$. If g is consistent with Q^-, then there is a $c' \in g$ falsified by two different $x^{[c_i]}, x^{[c_j]} \in Q^-$ since we have $m+1$ assignments in Q^- but strictly fewer clauses in g. Lemma 1 guarantees that $x^{[c_i]} \cap x^{[c_j]} \not\models c'$ and therefore g is falsified by $x^{[c_i]} \cap x^{[c_j]}$ as well. But $x^{[c_i]} \cap x^{[c_j]} \in Q_2^+$ and satisfies f. We conclude that no g can be consistent with f over Q_2^+. $\square$

By duality of the boolean operators and DNF vs. CNF representations we get

Corollary 7 *The classes monotone DNF,anti-monotone DNF, monotone CNF, anti-monotone CNF have certificates of size* $\min\{(m+1)n, \binom{m+1}{2} + m + 1\}$.

Constructing certificates for unate expressions appears harder at first since there are many more g functions that may be consistent with Q_1 or Q_2. Nonetheless essentially the same construction works here as well. Since for unate classes we define an orientation to transform the function to be monotone rather than anti-monotone, one would need the dual of the previous construction. To make the notation similar to the previous case we present the result for DNF which means taking the dual again so that we can use intersection as before.

Theorem 8 *Unate DNFs have polynomial size certificates with* $p(m,n) = m$ *and* $q(m,n) = \min\{(m+1)n, \binom{m+1}{2} + m + 1\}$.

Proof: Fix $m, n > 0$. Fix any $f \subseteq \{0,1\}^n$ s.t. $|f|_{unateDNF} > p(m,n) = m$. Now we proceed by cases.

Case 1. f is not unate. In this case, there must exist four assignments $x, y, z, w \in$

$\{0,1\}^n$ and a position i $(1 \leq i \leq n)$ such that:

- $x[j] = y[j]$ for all $1 \leq j \leq n, j \neq i$ and $x[i] < y[i]$
- $z[j] = w[j]$ for all $1 \leq j \leq n, j \neq i$ and $z[i] > w[i]$
- $f(x) > f(y)$ and $f(z) > f(w)$

Let $Q = \{x, y, z, w\}$. Notice that $|Q| \leq q(m,n)$. To see that no unate DNF can be consistent with f over Q, take any unate DNF g and suppose it is consistent. Let b be an orientation for g. If $b[i] = 0$ then we have that $x \leq_b y$ but $g(x) > g(y)$. If $b[i] = 1$ then $z \leq_b w$ but $g(z) > g(w)$. Therefore there cannot be any unate function consistent with f over Q.

Case 2. f is unate. Let a be any orientation showing that f is unate. Suppose w.l.o.g. (just renumber variables accordingly) that $a = 0^r 1^{n-r}$ where r is the number of monotone variables in f. Suppose that the variables in f are $\{v_1, ..., v_n\}$ and consider any minimal DNF representation $t_1 \vee t_2 \vee ... \vee t_m \vee ... \vee t_k$ of f. Notice that $k \geq m+1$ since $|f|_{unateDNF} > p(m,n) = m$. Since a is an orientation for f, and the DNF is minimal non-redundant the variables $\{v_1, ..., v_r\}$ appear always positive in the DNF and variables $\{v_{r+1}, ..., v_n\}$ appear always negated. Define j-th value of assignment $x^{[t_i]}$ as (for $1 \leq j \leq n$):

$$x^{[t_i]}[j] = \begin{cases} 1 \text{ if } j \leq r \text{ and } v_j \text{ appears in } t_i \\ 0 \text{ if } j \leq r \text{ and } v_j \text{ does not appear in } t_i \\ 0 \text{ if } j > r \text{ and } \bar{v}_j \text{ appears in } t_i \\ 1 \text{ if } j > r \text{ and } \bar{v}_j \text{ does not appear in } t_i \end{cases}$$

Notice that if f does not depend on a variable v_j, so that it does not appear in any of the terms, then it has the same value in all the assignments.

Let 0_j be defined as above. For the first construction let $Q_1 = Q^+ \cup Q_1^-$ where

$$Q^+ = \left\{ x^{[t_i]} \,\middle|\, 1 \leq i \leq m+1 \right\} \text{ and}$$

$$Q_1^- = \left\{ x^{[t_i]} \cap_a (a \oplus 0_j) \,\middle|\, 1 \leq i \leq m+1 \text{ and } x^{[t_i]}[j] = 1 - a[j]. \right\}$$

Notice that $a \oplus 0_j$ has all bits except the jth at their maximal value so $x^{[t_i]} \cap_a (a \oplus 0_j)$ flips the jth bit in $x^{[t_i]}$ to its minimum value. Each relevant variable has at least one pair of assignments in Q^+, Q_1^- with Hamming distance 1 showing the direction of its influence. Therefore any unate g consistent with Q_1 must have all variable polarities set correctly. As a result, the argument for the monotone case shows that any unate g with at most m terms over the relevant variables cannot be consistent with f over Q_1. Since irrelevant variables have a constant value in Q_1 they cannot affect consistency of any potential g.

For the second construction let $Q_2 = Q^+ \cup Q_2^-$ where Q^+ is defined as before and

$$Q_2^- = \left\{ x^{[t_i]} \cap_a x^{[t_j]} \;\middle|\; 1 \leq i < j \leq m+1 \right\}.$$

As before it is easy to see that the assignments in Q^+ are positive and assignments in Q_2^- are negative for f.

It is left to show that no unate DNF g s.t. $|g|_{unateDNF} \leq m$ is consistent with f over Q_2. If g is consistent with Q^+, then there is a $t' \in g$ satisfied by two assignments $x^{[t_i]}, x^{[t_j]} \in Q^+$. By Lemma 1 we get that $x^{[t_i]} \cap_a x^{[t_j]} \models t'$ and so it satisfies g as well. Since $x^{[t_i]} \cap_a x^{[t_j]} \in Q_2^-$ and it falsifies f, g is not consistent with f over Q_2. $\square$

Corollary 9 *The class of unate CNF has polynomial size certificates with* $p(m,n) = m$ *and* $q(m,n) = \min\{(m+1)n, \binom{m+1}{2} + m + 1\}$.

4 Certificates for Horn CNF

For anti-monotone CNF we could use the assignments defined by the clauses to generate the certificate. In particular the property from Remark 6 shows that each such assignment falsifies the clause generating it but no other clause in the representation. Any non-redundant CNF representation has such a set of assignments (since otherwise some clause is not needed in the representation) but it is not necessarily easy to find such assignments. As the following lemma shows for Horn expressions we can do this efficiently:

Lemma 10 *Let f be a non-redundant Horn CNF. For every clause c in f, we can efficiently find an assignment $x^{[c]}$ s.t. $x^{[c]}$ falsifies c but satisfies every other clause in f.*

Proof: Note first that such an assignment must exist since $f \setminus c \not\models c$ implies that there is an x such that $x \models f \setminus c$ and $x \not\models c$. Now since $f \setminus c$ is Horn and using Lemma 1 we see that if x, y are two different assignments satisfying this condition then so is $x \cap y$. So there is a unique minimal assignment satisfying this property. The minimal assignment can be found by finding the minimal model of $(f \setminus c) \wedge \overline{c}$. $\square$

Remark 11 While the previous lemma shows how to find the assignments efficiently they are not as explicitly related to the syntax of the representation as in the monotone case. It is interesting to note that given any non-redundant Horn CNF we can "saturate" it by adding implied propositions to the antecedents of rules. For example, if $f = (a \rightarrow b) \wedge (a \rightarrow c)$ we change the representation to $f = (a \rightarrow b) \wedge (ab \rightarrow c)$. One can show that if this is done sequentially until no more changes can be made then the final representation

has a syntactic property as in Remark 6. This construction was used in a previous version of this paper [27] and the improvement in Lemma 10 was suggested by an anonymous referee.

Theorem 12 *Horn CNFs have polynomial size certificates with* $p(m,n) = m(n+1)$ *and* $q(m,n) = \binom{m+1}{2} + m + 1$.

Proof: Fix $m, n > 0$. Fix any $f \subseteq \{0,1\}^n$ s.t. $|f|_{hornCNF} > p(m,n) = m(n+1)$. Again, we proceed by cases.

Case 1. f is not Horn. By Eq. (4), there must exist two assignments $x, y \in \{0,1\}^n$ s.t. $x \models f$ and $y \models f$ but $x \cap y \not\models f$. Let $Q = \{x, y, x \cap y\}$. Again by Eq. (4) no Horn CNF can be consistent with Q. Moreover, $|Q| = 3 \leq q(m,n)$.

Case 2. f is Horn. Let $c_1 \wedge c_2 \wedge .. \wedge c_{k'}$ be a minimal non-redundant representation of f. Notice that $k' \geq m(n+1)+1$ since $|f|_{hornCNF} > p(m,n) = m(n+1)$. Since there are more than $m(n+1)$ clauses, there must be at least $m+1$ clauses sharing a single consequent in f (there are at most $n+1$ different consequents among the clauses in f, including the constant `false`). Let these clauses be $c_1 = s_1 \rightarrow b, \ldots, c_k = s_k \rightarrow b$, with $k \geq m+1$. Let $x^{[c_i]}$ be the assignment that satisfies the conditions of Lemma 10 for c_i. Let $Q = Q^+ \cup Q^-$ where

$$Q^- = \left\{ x^{[c_i]} \;\middle|\; 1 \leq i \leq m+1 \right\} \text{ and}$$

$$Q^+ = \left\{ x^{[c_i]} \cap x^{[c_j]} \;\middle|\; 1 \leq i < j \leq m+1 \right\}.$$

Notice that $|Q| = |Q^+| + |Q^-| \leq \binom{m+1}{2} + m + 1 = q(m,n)$. The assignments in Q^- are negative for f by Lemma 10. We next show that every assignment in Q^+ satisfies every clause in f and therefore also satisfies f. Take any assignment $x^{[c_i]} \cap x^{[c_j]} \in Q^+$. For clauses c other than c_i and c_j, Lemma 10 guarantees that $x^{[c_i]} \models c$ and $x^{[c_j]} \models c$ and therefore $x^{[c_i]} \cap x^{[c_j]} \models c$ since c is Horn. To see that $x^{[c_i]} \cap x^{[c_j]} \models c_i$, suppose by way of contradiction that it does not. Since both $x^{[c_j]}$ and $x^{[c_i]}$ have the bit corresponding to their consequent set to 0 by construction (c_i and c_j share the same consequent), it must be that $x^{[c_i]} \cap x^{[c_j]}$ satisfies the antecedent of c_i. Therefore $x^{[c_j]}$ must also satisfy the antecedent of c_i, and $x^{[c_j]} \not\models c_i$ in contradiction with Lemma 10. We can prove the remaining case $x^{[c_i]} \cap x^{[c_j]} \models c_j$ analogously.

The argument that no Horn CNF g s.t. $|g|_{hornCNF} \leq m$ is consistent with f over Q is analogous to the anti-monotone case. □

Remark 13 The construction above relies on the fact that we can find many clauses with the same consequent. This fact does not hold in first order logic since the number of possible consequents is not bounded and therefore this hinders generalization. It is thus worth noting that a related construction with

slightly worse bounds does not require identical consequents. Consider again the minimal representation $c_1 \wedge c_2 \wedge \ldots \wedge c_{k'}$ with $k' \geq m(n+1)+1$. In this construction we use a larger Q^-

$$Q^- = \left\{ x^{[c_i]} \;\middle|\; 1 \leq i \leq m(n+1)+1 \right\}$$

and in addition we use the set

$$Q_{\text{others}} = \left\{ x^{[c_i]} \cap x^{[c_j]} \;\middle|\; 1 \leq i < j \leq m(n+1)+1 \right\}.$$

Note that assignments in Q_{others} may be either positive or negative since we have not restricted the consequent of clauses in Q^-. However, since Q^- is large, we get that some clause of g captures at least $n+2$ assignments in Q^-. We now consider the use of assignments from saturated expressions, and consider the relation between antecedents of different clauses generating these $n+2$ assignments. Since subsumption chains for antecedents (given by the subset relation over variables) are of length at most $n+1$, any set of clauses of this size must have a pair of clauses whose antecedents do not subsume one another. As a result there is at least one pair of clauses with incomparable antecedents, so that the intersection of assignments satisfies f but falsifies g so that g is not consistent with f over the certificate set. Unfortunately, subsumption chains for antecedents in first order logic can be long [28] so there are still obstacles in lifting the construction.

5 Learning from entailment

Work in inductive logic programming addresses learning formulas in first order logic and several setups for representing examples have been studied. The setup studied above where an example is an assignment in propositional logic generalizes to using first order structures (also known as interpretations) as examples. The model is therefore known as learning from interpretations [29]. In the model of learning from entailment an example is a clause. A clause example is positive if it is implied by the target and negative otherwise. Therefore a certificate in this context is a set of clauses. In particular, as in the previous case, for any expression f whose size is more than $p(m,n)$, a set Q of at most $q(m,n)$ clauses must satisfy that for any $g \in \mathcal{B}_m$ at least one element c of Q separates f and g, that is $f \models c$ and $g \not\models c$ or vice versa. We present a general transformation that allows us to obtain an entailment certificate from an interpretation certificate. Similar observations have been made before in different contexts, e.g. [30,20], where one transforms efficient algorithms instead of just certificates.

Definition 14 *Let x be an assignment. Then $ones(x)$ is the set of variables that are set to 1 in x. We slightly abuse notation and write $ones(x)$ to denote*

also the conjunction of the variables in the set $ones(x)$.

Lemma 15 *Let* f *be a boolean expression and* x *an assignment. Then,*

$$x \models f \text{ if and only if } f \not\models (ones(x) \rightarrow \bigvee_{b \notin ones(x)} b).$$

Proof: Suppose $x \models f$. Suppose by way of contradiction that $f \models (ones(x) \rightarrow \bigvee_{b \notin ones(x)} b)$. But since $x \not\models (ones(x) \rightarrow \bigvee_{b \notin ones(x)} b)$ we conclude that $x \not\models f$, which contradicts our initial assumption. Now, suppose $x \not\models f$. Hence, there is a clause $s \rightarrow \bigvee_i b_i$ in f falsified by x. This can happen only if $s \subseteq ones(x)$ and $b_i \notin ones(x)$ for all i. Clearly, $(s \rightarrow \bigvee_i b_i) \models (ones(x) \rightarrow \bigvee_{b \notin ones(x)} b)$. Therefore $f \models (ones(x) \rightarrow \bigvee_{b \notin ones(x)} b)$. □

Theorem 16 *Let* S *be an interpretation certificate for an expression* f *w.r.t. a class* $\mathcal{B}$ *of boolean expressions. Then, the set* $\{ones(x) \rightarrow \bigvee_{b \notin ones(x)} b \mid x \in S\}$ *is an entailment certificate for* f *w.r.t.* $\mathcal{B}$.

Proof: If S is an interpretation certificate for f w.r.t. some class $\mathcal{B}$ of propositional expressions, then for all $g \in \mathcal{B}$ there is some assignment $x \in S$ such that $x \models f$ and $x \not\models g$ or vice versa. Therefore, by Lemma 15, it follows that $f \not\models (ones(x) \rightarrow \bigvee_{b \notin ones(x)} b)$ and $g \models (ones(x) \rightarrow \bigvee_{b \notin ones(x)} b)$ or vice versa. Given the arbitrary nature of g the theorem follows. □

Remark 17 In the theorem above we include non-Horn clauses in the certificate. This is necessary since otherwise one cannot distinguish a function f from its Horn least upper bound [26,31], the function that is equivalent to the conjunction of all Horn clauses implied by f. For example, one cannot distinguish $f = \{a \rightarrow b, b \rightarrow c \vee d\}$ from $g = \{a \rightarrow b\}$ with Horn clauses only. It is worth noting, however, that a learning algorithm can use these certificates while making queries on Horn clauses only. The algorithm in [15,14] simulates the Halving Algorithm. In this process the algorithm constructs various functions f and asks membership queries on the examples in their certificates, i.e. in our case on the clauses. For a Horn expression T it holds that $T \models s \rightarrow b_1 \vee ... \vee b_k$ if and only if $T \models s \rightarrow b_i$ for some i. Thus, instead of asking a membership query on $s \rightarrow b_1 \vee ... \vee b_k$, the algorithm can ask k membership queries on $s \rightarrow b_i$ and reconstruct the answer. So while the certificate must include non-Horn clauses, the queries can avoid those.

6 Certificate size lower bounds

The certificate results above imply that unate and Horn CNF are learnable with a polynomial number of queries but as mentioned above this was already known. It is therefore useful to review the relationship between the certificate

size of a class and its query complexity. Recall from Theorem 4 that we have $CS(\mathcal{B}, n, m, p(n,m)) \leq QC(\mathcal{B}, n, m, p(n,m)) \leq CS(\mathcal{B}, n, m, p(n,m)) \log(|\mathcal{B}_m|)$. We note first that positive certificate results are not likely to improve known upper bounds for these classes. For the class of monotone DNF there is an algorithm that achieves query complexity $O(mn)$ [2,1]. In this case we have $\log(|monotoneDNF_m|) = \Theta(mn)$, so a certificate result is not likely to improve the known learning complexity. In the case of Horn CNF, there is an algorithm that achieves query complexity $O(m^2 n)$ [4]. Since again $\log(|HornCNF_m|) = \Theta(mn)$ improving on the known complexity $O(m^2 n)$ would require a certificate for Horn of size $o(m)$.

The results in this section show that this is not possible and in fact that our certificate constructions are optimal. We do this by giving lower bounds on certificate size. Naturally, these also imply lower bounds for the learning complexity.

In particular, for every m, n with $m < n$ we construct an n-variable monotone DNF f of size $\leq n$ and show that any certificate that f has more than m terms must have cardinality at least $q(m,n) = m + 1 + \binom{m+1}{2}$. This construction is shown for $p(n,m) = m$ thus giving lower bounds for strongly proper learning the class. We also give a variant where the size of f is n and where $m < n$ can be chosen arbitrarily. Thus the lower bound on learning complexity holds for any hypothesis expansion $p(n,m) < n$. For $m > n$ we show that there is a monotone DNF of size $m+1$ that requires a certificate of size $\Omega(mn)$. Again the bound is tight for strongly proper learning of monotone expressions. The lower bounds apply for Horn expressions as well where for $m > n$ we have a gap between $O(m^2)$ upper bound and $\Omega(mn)$ lower bound. The result for $m < n$ is given in the next two theorems:

Theorem 18 *Any certificate construction for monotone DNF for $m < n$ with $p(m,n) = m$ has size $q(m,n) \geq m + 1 + \binom{m+1}{2}$.*

Proof: Let $X_n = \{x_1, \ldots, x_n\}$ be the set of n variables and let $m < n$. Let $f = t_1 \vee \cdots \vee t_{m+1}$ where t_i is the term containing all variables (unnegated) except x_i. Such a representation is minimal and hence f has size exactly $m+1$. We show that for any set Q of size less than $m+1+\binom{m+1}{2}$ there is a monotone DNF with at most m terms consistent with f over Q.

If Q contains at most m positive assignments of weight $n-1$ then it is easy to see that the function with minterms corresponding to these positive assignments is consistent with f over Q. Hence we may assume that Q contains at least $m+1$ positive assignments of weight $n-1$. Thus if $|Q| < m+1+\binom{m+1}{2}$ then Q must contain strictly less than $\binom{m+1}{2}$ negative assignments. Notice that all the intersections between pairs of positive assignments of weight $n-1$ are different and there are $\binom{m+1}{2}$ such intersections. It follows that Q must

be missing some intersection between some pair of positive assignments in Q. But then there is an m-term monotone DNF consistent with Q which uses one term for the missing intersection and $m-1$ terms for the other $m-1$ positive assignments. $\Box$

The next theorem improves the hypothesis expansion from $p(n,m) = m$ to any choice satisfying $p(n,m) < n$.

Theorem 19 *Any certificate construction for monotone DNF for $m < n$ with $p(m,n) < n$ has size $q(m,n) \geq m + 1 + \binom{m+1}{2}$.*

Proof: Let $q(m,n) = m + 1 + \binom{m+1}{2}$ and define $f = \bigvee_{i \in \{1,..,n\}} t_i$ where t_i is the term containing all variables (unnegated) except x_i. Clearly, all t_i are minterms, f has size exactly n and f is monotone. We show that for any $m < n$ and any set of assignments Q of cardinality strictly less than $q(m,n)$, there is a monotone function g of at most m terms consistent with f over Q.

We first argue that w.l.o.g. we can assume that all the assignments in the potential certificate Q have weight $n-1$ (positive assignments) or weight $n-2$ (negative assignments). If Q contains the positive assignment 1^n, then we replace it by any assignment that of weight $n-1$. If Q contains a negative assignment x of weight smaller than $n-2$, then we replace it by any assignment $x' \geq x$ of weight $n-2$. Let Q' be the set obtained by replacing all these assignments of weight exactly n or smaller than $n-2$ in the manner described. Now any monotone function g consistent with Q' is also consistent with Q. As a result if Q' is not a certificate then neither is Q.

We next show that if $|Q| < q(m,n)$ then there exists a function g consistent with Q. Now since assignments in Q have weight $n-1$ or $n-2$ we can model the problem of finding a suitable monotone function as a graph coloring problem. We map Q into a graph $G_Q = (V,E)$ where $V = \{p \in Q \mid f(p) = 1\}$ and $E = \{(p_1,p_2) \mid \{p_1,p_2,p_1 \cap p_2\} \subseteq Q\}$. Let $|V| = v$ and $|E| = e$.

First we show that if G_Q is m-colorable then there is a monotone function g of DNF size at most m that is consistent with f over Q. It is sufficient that for each color c we find a term t_c that (1) is satisfied by the positive assignments in Q that have been assigned color c, with the additional condition that (2) t_c is not satisfied by any of the negative assignments in Q. We define t_c as the minterm corresponding to the intersection of all the assignments colored c by the m-coloring. Property (1) is clearly satisfied, since no variable set to zero in any of the assignments is present in t_c. To see that (2) holds it suffices to notice that the assignments colored c form an independent set in G_Q and therefore none of their pair-wise intersections is in Q. By the assumption no negative point below the intersections is in Q either. The resulting consistent function g contains all minterms t_c. Since the graph is m-colorable, g has at

most m terms.

It remains to show that G_Q is m-colorable. Note that the condition $|Q| < q(m,n)$ translates into $v + e < q(m,n)$ in G_Q. If $v \leq m$ then there is a trivial m-coloring. For $v \geq m+1$, we have $e < \binom{m+1}{2} - 1$ so it suffices to prove the following lemma to complete the proof of Theorem 19:

Lemma 20 *Any v-node graph with $v \geq m+1$ with at most $\binom{m+1}{2} - 1$ edges is m colorable.*

We prove this lemma by induction on v. The base case is $v = m+1$; in this case since the graph has at most $\binom{m+1}{2} - 1$ edges it can be colored with only m colors by reusing one color for the missing edge. For the inductive step, note that any v-node graph which has at most $\binom{m+1}{2} - 1$ edges must have some node with fewer than m neighbors since otherwise there would be at least $vm/2 \geq \frac{(m+2)m}{2} = \frac{(m+1)m}{2} + \frac{m}{2} > \binom{m+1}{2} - 1$ edges in the graph. By the induction hypothesis there is an m-coloring of the $(v-1)$-node graph obtained by removing this node of minimum degree and its incident edges. But since the degree of this node was less than m in G, we can color G using at most m colors. This concludes the proof of Lemma 20 and of Theorem 19. □

Finally, we give an $\Omega(mn)$ lower bound on certificate size for monotone DNF for the case $m > n$. Like Theorem 18 this result gives a lower bound on query complexity for any strongly proper learning algorithm.

Theorem 21 *Any certificate construction for monotone DNF for $m > n$ with $p(m,n) = m$ has size $q(m,n) = \Omega(mn)$.*

Proof: Fix any constant k. We show that for all n and for all $m = \binom{n}{k} - 1$, there is a function f of monotone DNF size $m+1$ such that any certificate showing that f has more than m terms must contain $\Omega(nm)$ assignments.

We define f as the function whose satisfying assignments have at least $n-k$ bits set to 1. Notice that the DNF size of f is exactly $\binom{n}{k} = m+1$. Let P be the set of assignments corresponding to the minterms of f, i.e. P consists of all assignments that have exactly $n-k$ bits set to 1. Let N be the set of assignments that have exactly $n-(k+1)$ bits set to 1. Notice that f is positive for the assignments in P but negative for those in N. Clearly, assignments in P are minimal weight positive assignments and assignments in N are maximal weight negative assignments. Note that $|P| = \binom{n}{k}$ and $|N| = (m+1)\frac{n-k}{k+1} = \binom{n}{k+1} = \Omega(mn)$ for constant k. Moreover, any assignment in N is the intersection of two assignments in P.

We next show that any certificate for f must have size at least $|P| + |N|$. As in the previous proof, we may assume w.l.o.g. that any certificate Q contains

assignments in $P \cup N$ only. Let $Q \subset P \cup N$. If Q has at most m positive assignments then it is easy to construct a function consistent with Q regardless of how negative examples are placed. Otherwise, Q contains all the $m+1$ positive assignments in P and the rest are assignments in N. If Q misses any assignment in N then we build a consistent function by using the minterm corresponding to the missing intersection to "cover" two of the positive assignments with just one term. The remaining $m-1$ positive assignments in P are covered by one minterm each. Hence, any certificate Q must contain $P \cup N$ and thus is of size $\Omega(nm)$. $\Box$

Finally, we observe that all the lower bounds above apply to unate and Horn CNF expressions as well. This follows from the fact that the function f used in the construction is outside the class (has size more than m in all cases) and that the function g constructed is in the class (since monotone DNF is a special case of unate DNF and Horn DNF). We therefore have:

Corollary 22 *Any certificate construction for unate CNF (DNF) and for Horn CNF (DNF) must satisfy the bounds given in Theorems 18, 19 and 21.*

7 An exponential lower bound for renamable Horn

In this section we show that renamable Horn CNF expressions do not have polynomial certificates. This answers an open question of [22] and implies that the class of renamable Horn CNF is not exactly learnable using a polynomial number of membership and equivalence queries. In the following let $\mathcal{B}$ be the class of renamable Horn expressions.

To show non-existence of polynomial certificates, we need to prove the following: for all two-variable polynomials $p(\cdot,\cdot)$ and $q(\cdot,\cdot)$ there exist $n, m > 0$ and a boolean function $\hat{f} \subseteq \{0,1\}^n$ with $|\hat{f}|_{\mathcal{B}} > p(m,n)$ such that for every $Q \subseteq \{0,1\}^n$, either (1) $|Q| > q(m,n)$ or (2) some $g \in \mathcal{B}_m$ is consistent with f over Q.

In order to show this, we define a function $\hat{f}$ that is not renamable Horn, so that $|\hat{f}|_{\mathcal{B}} = \infty > p(m,n)$ holds for any function $p(m,n)$ and the requirement can be simplified. Utilizing this simplification, what we show is: for each n which is a multiple of 3, there exists a non-renamable Horn $\hat{f} \subseteq \{0,1\}^n$ s.t. if no $g \in \mathcal{B}_{n^6}$ is consistent with $\hat{f}$ over some set of assignments Q (i.e. we are taking $m = n^6$), then $|Q| \geq \frac{1}{3}2^{2n/3}$. Equivalently, for every such n every certificate Q that $\hat{f}$ is not a renamable Horn CNF function of size n^6 has to be of exponential size. This is clearly sufficient to prove the non-existence of polynomial certificates for renamable Horn boolean functions.

We say that a set Q such that no $g \in \mathcal{B}_{n^6}$ is consistent with $\hat{f}$ over Q is a *certificate that $\hat{f}$ is not small renamable Horn.* The following lemma is useful:

Lemma 23 *Let f be a satisfiable renamable Horn function. Then there is an orientation c for f such that $c \models f$.*

Proof: Let c' be an orientation of f such that $c' \not\models f$. Let c be the positive assignment of f which is minimal with respect to the partial order imposed by $\leq_{c'}$. There exists a single such assignment. This can be seen via Eq. (5) since if a and b are both positive assignments unrelated in the partial order imposed by $\leq_{c'}$, then $c'' = a \cap_{c'} b$ is positive.

We claim that c is an orientation for f. It suffices to show $a \cap_{c'} b = a \cap_c b$ for all positive assignments a and b. We show that $(a \cap_{c'} b)[i] = (a \cap_c b)[i]$ for all $1 \leq i \leq n$. If i is such that $c[i] = c'[i]$ then clearly $(a \cap_{c'} b)[i] = (a \cap_c b)[i]$. Let i be such that $c[i] \neq c'[i]$. Then every positive assignment sets the bit i like $c[i]$: if $a[i] \neq c[i]$ then $(a \cap_{c'} c)[i] = c'[i]$ and thus $(a \cap_{c'} c) <_{c'} c$ (strictly), contradicting the minimality of c. Thus $a[i] = b[i] = c[i]$ and $(a \wedge b)[i] = (a \vee b)[i]$, and therefore $(a \cap_c b)[i] = (a \cap_{c'} b)[i]$. $\square$

We next define the function $\hat{f}$. As the next lemma shows $\hat{f}$ is not renamable Horn. The function has two useful properties: it has a very small number of satisfying assignments, and the hamming distance between these is large. The second property helps guarantee that the certificate is large. The first property is used to bound the size of the hypothesis expansion. This is the main difference from an earlier result of [22] where a weaker type of lower bound was proved. In that result Feigelson [22] gave a class of functions and showed that a superpolynomial size set of assignments is needed to certify that they are not renamable Horn. However, a certificate only needs to certify that the function is not *small* renamable Horn so the result did not have direct implications for certificate size. This is addressed by the current construction.

Definition 24 *Let $n = 3k$ for some $k \geq 1$. We define $\hat{f} : \{0,1\}^n \to \{0,1\}$ to be the function whose only satisfying assignments are $0^k1^k1^k, 1^k0^k1^k$, and $1^k1^k0^k$.*

Lemma 25 *The function $\hat{f}$ defined above is not renamable Horn.*

Proof: To see that a function f is not renamable Horn with orientation c it suffices to find a triple (p_1, p_2, q) such that $p_1 \models f$, $p_2 \models f$ but $q \not\models f$ where $q = p_1 \cap_c p_2$. By Lemma 23 it is sufficient to check that the three positive assignments are not valid orientations for f:
The triple $(1^k1^k0^k, 1^k0^k1^k, 1^k1^k1^k)$ rejects $c = 0^k1^k1^k$.
The triple $(0^k1^k1^k, 1^k1^k0^k, 1^k1^k1^k)$ rejects $c = 1^k0^k1^k$.
The triple $(0^k1^k1^k, 1^k0^k1^k, 1^k1^k1^k)$ rejects $c = 1^k1^k0^k$. $\square$

Consider next how certificates for renamable Horn may be structured. If an orientation c does not witness that f is renamable Horn then there is a triple of assignments (p_1, p_2, q) such that $p_1 \models f$, $p_2 \models f$, $q = p_1 \cap_c p_2$ but $q \not\models f$. However, a certificate does not necessarily need to have such a triple explicitly. To illustrate this consider three positive assignments x, y, z and a negative assignment w such that $w = (x \wedge_c y) \wedge_c z$. Clearly $\{x, y, z, w\}$ show that c is not an orientation for f but the assignment $(x \wedge_c y)$ is not in the set and there is no explicit triple of this form. As the next lemma shows a weaker notion of triple must appear in any certificate.

We say that a triple (p_1, p_2, q) such that $p_1 \models f$, $p_2 \models f$ but $q \not\models f$ is *suitable* for c if $q \leq_c p_1 \cap_c p_2$.

Lemma 26 *If Q is a certificate that $\hat{f}$ is not small renamable Horn with orientation c, then Q includes a suitable triple (p_1, p_2, q) for c.*

Proof: Suppose that a certificate Q that $\hat{f}$ is not small renamable Horn with orientation c does not include any suitable triple (p_1, p_2, q) for c. We define a function g that is consistent with $\hat{f}$ on Q as follows:

$$g(x) = \begin{cases} 1 \text{ if } x \in Q \text{ and } x \models \hat{f} \\ 1 \text{ if } x \leq_c (s_1 \cap_c s_2) \text{ for any } s_1, s_2 \in Q \text{ s.t. } s_1 \models \hat{f} \text{ and } s_2 \models \hat{f} \\ 0 \text{ otherwise.} \end{cases}$$

The function g is consistent with Q since by assumption no negative example is covered by the second condition.

First we show that the function g is renamable Horn with orientation c. Consider any assignments p_1, p_2 that are positive for g, i.e., $p_1 \models g$ and $p_2 \models g$, and let $t = p_1 \cap_c p_2$. If p_1, p_2 are included in Q, then clearly $t \models g$ by the definition of g. If $p_1 \notin Q$ then $p_1 \leq_c (s_1 \cap_c s_2)$ for some positive $s_1, s_2 \in Q$ (second condition in the definition of g). Since $t \leq_c p_1 \leq_c (s_1 \cap_c s_2)$, then by the definition of g, $t \models g$ as well. The same reasoning applies for the remaining case $p_2 \notin Q$. Hence, g is renamable Horn with orientation c. Note that this part of the proof does not rely on specific properties of $\hat{f}$ and thus holds for any f which is not renamable Horn.

Now, we show that g is also *small*. We use the fact that our particular $\hat{f}$ is designed to have very few positive assignments. First notice that g only depends on the positive assignments in Q. Moreover, these must be positive assignments for $\hat{f}$. Suppose that Q contains any $l \leq 3$ of these positive assignments.

Let these be $x_1, \ldots, x_l$. A DNF representation for g is:

$$g = \bigvee_{1 \le i \le l} t_i \vee \bigvee_{1 \le i < j \le l} t_{i,j}$$

where t_i is the term that is true for the assignment x_i only and $t_{i,j}$ is the term that is true for the assignment $x_i \cap_c x_j$ and all assignments below it (w.r.t. c). Notice that we can represent this with just one term by removing literals that correspond to maximal values (w.r.t. c). For example, if $l = 2$ and $x_1 = 001111$, $x_2 = 110011$ and $c = 101001$ then $t_1 = \overline{v_1}\ \overline{v_2} v_3 v_4 v_5 v_6$, $x_1 \cap_c x_2 = 101011$, and the only variable at its maximal value is v_5 so $t_{1,2} = v_1 \overline{v_2} v_3 \overline{v_4} v_6$.

Since $l \le 3$, g has at most $3 + \binom{3}{2} = 6$ terms. Hence, g has CNF size at most n^6 (multiply out all terms to get the clauses). Note that so far we have shown that $|g|_{CNF} \le n^6$ but we must also show that $|g|_{Ren-Horn}$ is small. This follows from the well known fact that if a function h is Horn and g is a non-Horn CNF representation for h, then every clause in g can be replaced with a Horn clause which uses a subset of its literals; see e.g. [24] or Claim 6.3 of [26]. So the arbitrary CNF for g can be replaced with a renamable Horn CNF of the same size. We arrive at a contradiction: Q is not a certificate that $\hat{f}$ is not small renamable Horn with orientation c since $\tilde{g}$ is not rejected. □

Theorem 27 *For all $n = 3k$, there is a function $\hat{f} : \{0,1\}^n \to \{0,1\}$ which is not renamable Horn such that any certificate Q showing that the renamable Horn size of $\hat{f}$ is more than n^6 must have $|Q| \ge \frac{1}{3} 2^{2n/3}$.*

Proof: Recall that a triple (p_1, p_2, q) is suitable for c if $p_1 \models f$, $p_2 \models f$ but $q \not\models f$ where $q \le_c p_1 \cap_c p_2$. Consider any bit where p_1 and p_2 differ, that is $p_1[i] \ne p_2[i]$. In this case the intersection always obtains the minimal value $p_1[i] \cap_c p_2[i] = c[i]$. This also implies that $q[i] \le_c p_1[i] \cap_c p_2[i] = c[i]$ satisfies $q[i] = c[i]$. Now if p_1, p_2 have k bits with different values, any fixed q forces k bit values in c and therefore (p_1, p_2, q) is suitable for 2^{n-k} values of c.

Now we use the fact that the Hamming distance between any two positive assignments of $\hat{f}$ is $2n/3$. A negative example in Q can appear in at most 3 triples (only 3 choices for p_1, p_2), and hence any negative example in Q contributes to at most $3 \cdot 2^{n/3}$ orientations. The theorem follows since we need to reject all orientations. □

Corollary 28 *Renamable Horn CNFs do not have polynomial sized certificates.*

8 Conclusion

This paper provides a study of the certificate complexity of several well known representation classes for propositional expressions. Since certificates are known to characterize the query complexity of exact learning with queries our results have direct implications for learnability. In particular the paper provides certificate constructions and hence upper bounds on their size for monotone, unate and Horn expressions. Lower bounds for these classes are also derived and these are tight in some cases. An exponential lower bound for the class of renamable Horn expressions establishes that the class is not learnable with a polynomial number of queries. The following table summarizes the bounds obtained in this paper:

Class	*LowerBound*		*UpperBound*	
unate DNF/CNF $m < n$	$\binom{m+1}{2} + m + 1^*$	(Th. 19)	$\binom{m+1}{2} + m + 1$	(Th. 8)
unate DNF/CNF $m \geq n$	$\Omega(mn)^{**}$	(Th. 21)	$O(mn)$	(Th. 8)
Horn CNF $m < n$	$\binom{m+1}{2} + m + 1^*$	(Th. 19)	$\binom{m+1}{2} + m + 1$	(Th. 12)
Horn CNF $m \geq n$	$\Omega(mn)^{**}$	(Th. 21)	$\binom{m+1}{2} + m + 1$	(Th. 12)
renamable Horn CNF	$\frac{1}{3}2^{2n/3}$	(Th. 27)		

* For $p(m, n) < n$.

** Strong certificate size only, i.e. $p(m, n) = m$.

Several interesting questions remain unsolved. For Horn expressions with $m > n$ clauses there is a gap between the lower bound $\Omega(mn)$ and the upper bound $O(m^2)$. Also except for renamable Horn the lower bounds are for strongly proper learnability or a small expansion in hypothesis size $p(m, n) < n$. Identifying the certificate complexity and equivalently the query complexity of general DNF is an important open question. Finally, as mentioned in the introduction, there is an exponential gap between known lower bounds and upper bounds on learning complexity for first order Horn expressions. Certificates may provide a tool to resolve this gap and the constructions for the propositional special cases developed in this paper are natural starting points in such an endeavor.

Acknowledgments

We thank José Luis Balcázar for comments on earlier drafts, and Lisa Hellerstein for raising the question of certificate lower bound for strong learnability. We also thank the reviewers for helping to simplify some of the proofs and

improve the presentation.

References

[1] D. Angluin, Queries and concept learning, Machine Learning 2 (4) (1988) 319–342.

[2] L. G. Valiant, A theory of the learnable, Communications of the ACM 27 (11) (1984) 1134–1142.

[3] N. H. Bshouty, Simple learning algorithms using divide and conquer, in: Proceedings of the Conference on Computational Learning Theory, 1995, pp. 447–453.

[4] D. Angluin, M. Frazier, L. Pitt, Learning conjunctions of Horn clauses, Machine Learning 9 (1992) 147–164.

[5] M. Frazier, L. Pitt, Learning from entailment: An application to propositional Horn sentences, in: Proceedings of the International Conference on Machine Learning, Morgan Kaufmann, Amherst, MA, 1993, pp. 120–127.

[6] C. Reddy, P. Tadepalli, Learning Horn definitions with equivalence and membership queries, in: International Workshop on Inductive Logic Programming, Springer, Prague, Czech Republic, 1997, pp. 243–255, LNAI 1297.

[7] R. Khardon, Learning function free Horn expressions, Machine Learning 37 (1999) 241–275.

[8] M. Arias, R. Khardon, Learning closed Horn expressions, Information and Computation 178 (2002) 214–240.

[9] H. Arimura, Learning acyclic first-order Horn sentences from entailment, in: Proceedings of the International Conference on Algorithmic Learning Theory, Springer-Verlag, Sendai, Japan, 1997, pp. 432–445, LNAI 1316.

[10] C. Reddy, P. Tadepalli, Learning first order acyclic Horn programs from entailment, in: International Conference on Inductive Logic Programming, Springer, Madison, WI, 1998, pp. 23–37, LNAI 1446.

[11] K. Rao, A. Sattar, Learning from entailment of logic programs with local variables, in: Proceedings of the International Conference on Algorithmic Learning Theory, Springer-verlag, Otzenhausen, Germany, 1998, pp. 143–157, LNAI 1501.

[12] M. Arias, R. Khardon, Complexity parameters for first-order classes, in: Proceedings of the 13th International Conference on Inductive Logic Programming, Springer-Verlag, 2003, pp. 22–37, LNAI 2835. Full version to appear in the Machine Learning Journal.

[13] A. Blumer, A. Ehrenfeucht, D. Haussler, M. K. Warmuth, Learnability and the Vapnik-Chervonenkis dimension, Journal of the ACM 36 (4) (1989) 929–965.

[14] L. Hellerstein, K. Pillaipakkamnatt, V. Raghavan, D. Wilkins, How many queries are needed to learn?, Journal of the ACM 43 (5) (1996) 840–862.

[15] T. Hegedűs, On generalized teaching dimensions and the query complexity of learning, in: Proceedings of the Conference on Computational Learning Theory, ACM Press, New York, NY, USA, 1995, pp. 108–117.

[16] J. L. Balcázar, J. Castro, D. Guijarro, The consistency dimension and distribution-dependent learning from queries, in: Proceedings of the International Conference on Algorithmic Learning Theory, Springer, Tokyo, Japan, 1999, pp. 77–92, LNAI 1702.

[17] D. Angluin, Queries revisited, Theoretical Computer Science 313 (2004) 175–194.

[18] A. Feigelson, L. Hellerstein, Conjunctions of unate DNF formulas: Learning and structure, Information and Computation 140 (2) (1998) 203–228.

[19] L. Hellerstein, V. Raghavan, Exact learning of DNF formulas using DNF hypotheses, in: Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC-02), ACM Press, New York, 2002, pp. 465–473.

[20] L. De Raedt, Logical settings for concept learning, Artificial Intelligence 95 (1) (1997) 187–201, see also relevant Errata (forthcoming).

[21] A. del Val, On 2-SAT and renamable Horn, in: Proceedings of the National Conference on Artificial Intelligence, 2000, pp. 279–284.

[22] A. Feigelson, On boolean functions and their orientations: Learning, monotone dimension and certificates, Ph.D. thesis, Northwestern University, Evanston, IL, USA (Jun. 1998).

[23] M. Alekhnovich, M. Braverman, V. Feldman, A. Klivans, T. Pitassi, Learnability and automatizability, in: 45th Annual Symposium on Foundations of Computer Science (FOCS'04), IEEE Computer Society Press, Los Alamitos, 2004, pp. 621–630.

[24] J. C. C. McKinsey, The decision problem for some classes of sentences without quantifiers, J. Symbolic Logic 8 (1943) 61–76.

[25] A. Horn, On sentences which are true of direct unions of algebras, Journal of Symbolic Logic 16 (1956) 14–21.

[26] R. Khardon, D. Roth, Reasoning with models, Artificial Intelligence 87 (1–2) (1996) 187–213.

[27] M. Arias, R. Khardon, R. A. Servedio, Polynomial certificates for propositional classes, in: Proceedings of the Conference on Computational Learning Theory, Springer-Verlag, 2003, pp. 537–551, LNAI 2777.

[28] M. Arias, R. Khardon, The subsumption lattice and query learning, in: Proceedings of the International Conference on Algorithmic Learning Theory, Springer, Padova, Italy, 2004, pp. 410–424.

[29] L. De Raedt, S. Dzeroski, First order jk-clausal theories are PAC-learnable, Artificial Intelligence 70 (1994) 375–392.

[30] R. Khardon, D. Roth, Learning to reason with a restricted view, Machine Learning 35 (2) (1999) 95–117.

[31] B. Selman, H. Kautz, Knowledge compilation and theory approximation, J. of the ACM 43 (2) (1996) 193–224.